David’s Final Report

David Skrill

12/15/2020

Abstract

As Austin Texas is ranked as one of the best cities to live and raise a family in with great expansion in economy, education, outdoor activities, weather etc., the corresponding social issues such as economical segregation and racial profiling. In this study, we are seeking valid methods to evaluate the fairness in traffic stops. We found that

Contents

Introduction

Expanding economy, good schools, and many outdoor and indoor activities, Austin, Texas, seems to unite the best qualities to become the best city in the U.S. And the best city she has become. For four years in a row (2016 - 2020), Austin was ranked number one in “the best place to live” list by U.S News. However, like many other big cities, Austin has her own problems. The most prominent ones are economic segregation and racial profiling. In this project, our team will focus on the problem of traffic-stopping fairness. Notably, we aim to use the data sets provided by the Austin Police Department and data from the Stanford open policing project to help evaluate and expand our understanding of this problem in Austin. Our main report consists of three parts. First, we will do data exploratory analysis to get a big picture of Austin’s racial profiling issue. Second, we are going to fit statistical models to estimate the severity of this issue. Third, we propose a measure of fairness based on the difference of posterior hit rate among each race for each police officer.

Available Data

In this project, we utilized two main data sets. The primary data set is from the Stanford open policing project[1] (we call it Stanford data). This data set record stops made by the APD around ten years (2006.01.01 - 2016.06.30). This data set contains the date of the stops, the subject’s race, whether the person was searched or frisked, whether any contrabands were found, etc. However, it does not have the information about the specific time and locations the events occurred. Since we only half year of 2016 data, we decide to focus on 2006-2015 nine years of complete data with total 463,944 stop events recorded.

Our secondary data set is Racial Profiling dataset from the city of Austin website (we call it RP data). This data set contains similar information as in our first data set with additional information of the event time

[1]Stanford Open Policing Project (OPP): https://openpolicing.stanford.edu/data/

and locations as well as the race of the police available. Nevertheless, this data set misses a key element: the race of the drivers.

We also use US census websites to estimate the population and proportion of race in Austin. Specifically, we use 5-year average census population data for Stanford 2006-2016 racial profiling data, and 2019 census population data for 2019 Austin RP data. In addition, we also refer to the racial profiling reports from the Austin Police Department as an external resource to get the race information.

Summary Statistics

Stanford Data

Summary statistics for the Stanford data are as follows. The data covers about a ten year period (2006.01.01 - 2016.06.30). Not shown are unique officer identifiers.

% latex table generated in R 4.0.5 by xtable 1.8-4 package % Sun Apr 18 11:50:59 2021

	nobs	nmis	uniq	mean	SD	min	25%	50%	75%	max
subject_age	480091	3164	94	37.98	13.82	10.00	26.00	36.00	48.00	103.00
subject_sex	482881	374	2	0.30	0.46	0.00	0.00	0.00	1.00	1.00
frisk_performed	483255	0	2	0.02	0.15	0.00	0.00	0.00	0.00	1.00
search_conducted	483255	0	2	0.04	0.20	0.00	0.00	0.00	0.00	1.00
search_person	483255	0	2	0.03	0.18	0.00	0.00	0.00	0.00	1.00
search_vehicle	483255	0	2	0.02	0.15	0.00	0.00	0.00	0.00	1.00

% latex table generated in R 4.0.5 by xtable 1.8-4 package % Sun Apr 18 11:51:10 2021

	nobs	nmis	uniq	mean	SD	min	25%	50%	75%	max
contraband_found	19256	0	2	0.25	0.43	0.00	0.00	0.00	0.00	1.00
contraband_drugs	19256	0	2	0.01	0.12	0.00	0.00	0.00	0.00	1.00
contraband_weapons	19256	0	2	0.05	0.21	0.00	0.00	0.00	0.00	1.00
frisk_performed	19256	0	2	0.51	0.50	0.00	0.00	1.00	1.00	1.00

Table 1: Subject race.

Race	n	percent
asian/pacific islander	13167	0.0272466
black	72324	0.1496607
hispanic	123943	0.2564764
other	2626	0.0054340
unknown	3135	0.0064873
white	268058	0.5546950
NA	2	NA

Table 2: Search basis.

Search basis	n	percent
consent	3195	0.1659223
other	276	0.0143332
plain view	152	0.0078936
probable cause	15633	0.8118509
NA	463999	NA

Figure 1: Distribution of stops by unique officer ID

Investigating the Hit Rate

The "hit rate," defined here as the proportion of times an officer finds contraband given that a frisk has been performed, is a widely-used measure for assessing potentially-discriminatory policing. The hit rate can be thought of as a proxy for "evidence" when an officer decides whether to conduct a search or a frisk; a lower hit rate for a particular segment of the population can signal that an officer has a lower threshold of evidence when policing that population segment. In the following analysis, we examine the hit rate at the officer level. Because the analysis requires that officers have stopped all races under consideration, we restrict the analysis to only White, Black, and Hispanic subject races and to officers with 18 or more stops, corresponding to roughly the 90th percentile.

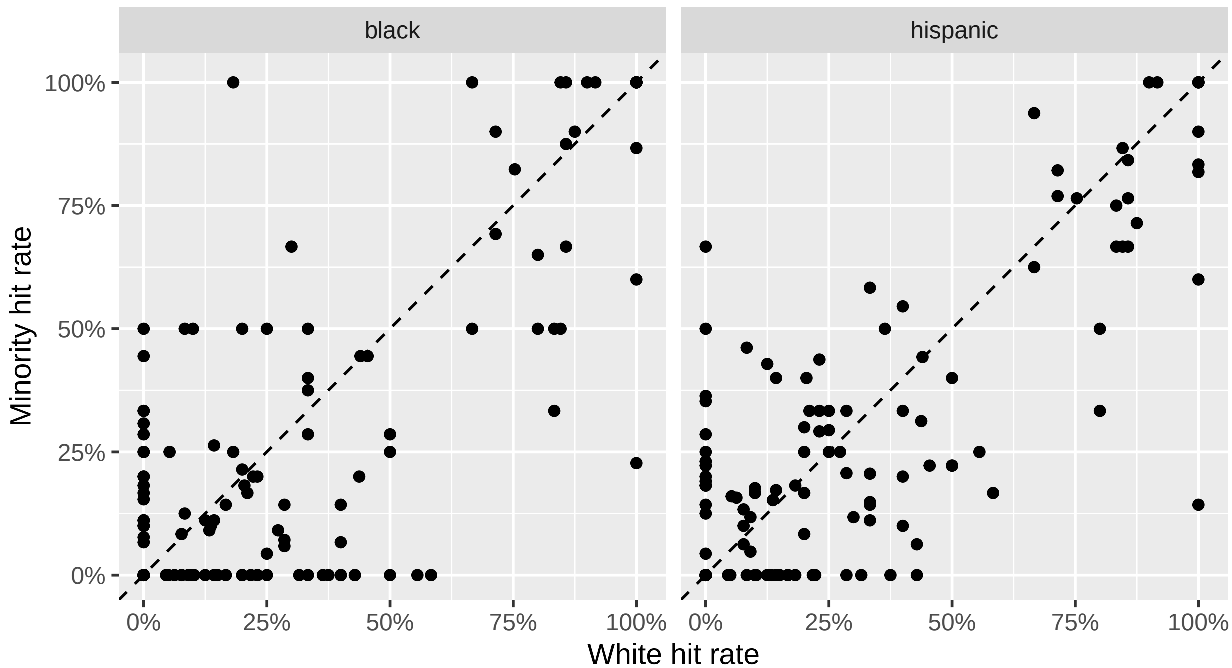


Figure 2: Hit rates for individual officers.

The above plot shows the hit rates for individual officers. An officer with an identical hit rate for white and minority subpopulations would be on the 45-degree line. Visually, it is difficult to determine a systematic trend, although it is clear that particular officers have hit rates that differ substantially by subpopulaton. It should be noted that the hit rate is highly variable with small sample sizes.

We proceed using a Bayesian hierarchical model. Under this model, we treat individual officers as belonging to a population of players and we seek to model both the hit rates of the officers and the variation of this population. This permits *partial pooling*, by which individual hit rates are biased towards the population average by an amount determined by the estimate of the population. For each officer, we consider three hit rates, one each for white, Black, and Hispanic subpopulations. We accomplish this by fitting separate logistic mixed effects models for each race, each with a weakly informative Normal prior on the log-odds with mean -1.2 and standard deviation 1.

Specifically, let θ_{jr} be the hit rate for for officer j and race r, y_{jr} be the number of hits, and K_{jr} the number of frisks. In the following, because we fit separate models, we assume for example $r = black$ and drop the r subscript. Assuming each officer's searches are independent Bernoulli trials

$$p(y_j|\theta_j) = \text{Binomial}(y_j|K_j, \theta_j)$$

We reparametrize the model in terms of the log-odds, α:

$$\alpha_j = \text{logit}(\theta_j) = \log \frac{\theta_j}{1-\theta_j}$$

We set a weakly informative prior centered at $\alpha_j = -1.3$, corresponding to $\theta_j \approx 0.2$. The model is therefore

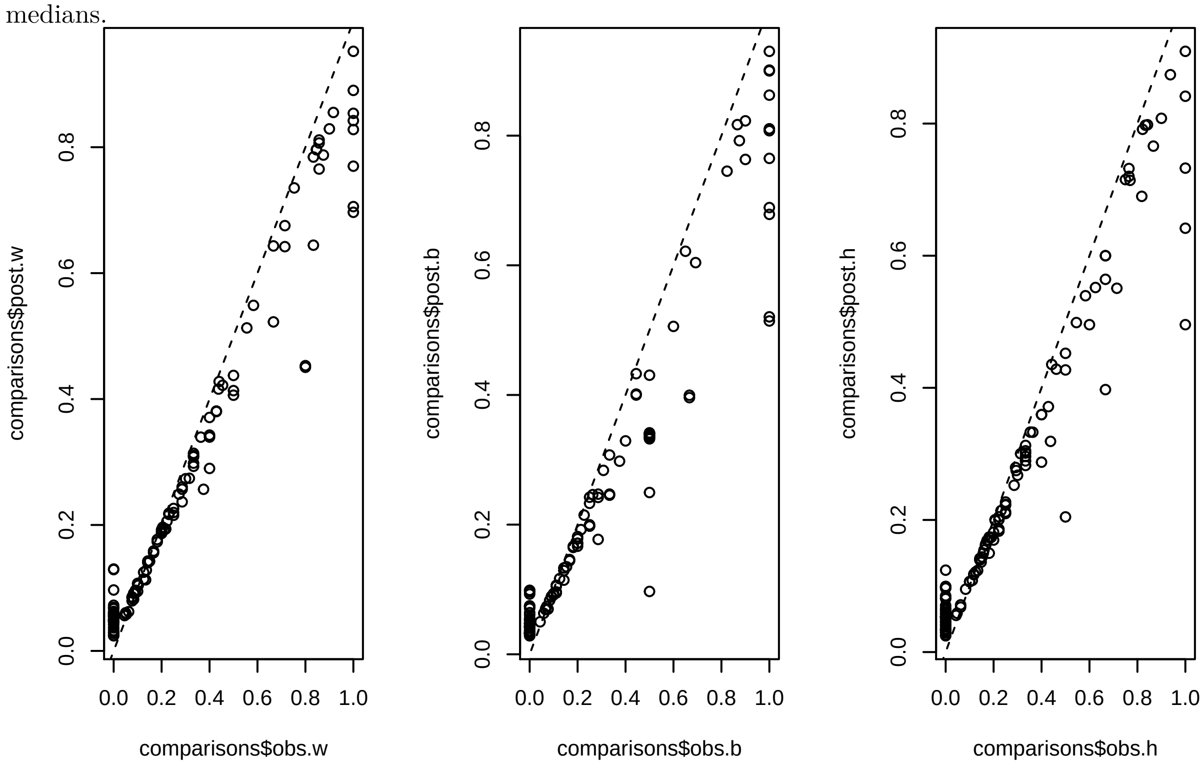
$$p(y_j|K_j, \alpha) = \text{Binomial}(y_j|K_j, \text{logit}^{-1}(\alpha_j))$$

We proceed using `stan_glmer` and the default prior on the covariance matrix. The result includes a posterior for each officer; we may transform from the log-odds back to hit rate to to obtain a posterior for the the hit rate for each officer. We model each race separately, and so obtain three posteriors for each officer.

Table 3: Posterior intervals for three races. From left to right: white, Black, Hispanic

x	2.5%	50%	97.5%	2.5%	50%	97.5%	2.5%	50%	97.5%
01db7098a7	0.003	0.055	0.317	0.021	0.198	0.617	0.065	0.214	0.450
020579eaad	0.021	0.095	0.248	0.002	0.033	0.203	0.047	0.162	0.359
02b0803fe3	0.003	0.056	0.324	0.111	0.242	0.420	0.037	0.174	0.423
0329f48f95	0.128	0.523	0.900	0.057	0.521	0.959	0.665	0.874	0.974
068ff01d47	0.054	0.188	0.415	0.032	0.332	0.833	0.082	0.268	0.556
	0.114	0.380	0.722	0.002	0.040	0.261	0.010	0.071	0.237

The the following, the effects of partial pooling are evident: the posterior medians aer baised towards the population average. Practically, this means that observed hit rates equal to zero have posterior medians that are small but positive, and perfect (or near-perfect) observed hit rates have somewhat smaller posterior medians.



Because part of this project is to “operationalize” fairness, we devised a measure by which the above posteriors can be converted into a rough “fairness score.” Because an officer that uses the same evidence threshold when deciding whether to frisk a subject regardless of race should have roughly equal hit rates for all three subpopulations, we reason that such an officer should have posterior medians that are close to each other for the three subpopulations. So, one can calculate a simple sum of squares statistic for each officer. Specifically, letting m_{jr} be the posterior median for officer j and race r, the sum of squares statistic S_j is

$$S_j = \sum_r (m_{jr} - \bar{m_j})^2$$

where $\bar{m_j}$ is the average of the three medians. Of course, this measure disregards all other information that could be gleaned from the posterior; an alternative might calculate the overlap between the posterior densities. However, we think this measure is relatively easy to understand and implement.

Table 4: Highest 5 scores.

ID	obs.w	obs.b	obs.h	post.w	post.b	post.h	count.w	count.b	count.h
1504c3bc16	1.000	0.227	0.143	0.697	0.215	0.142	2	22	7
3392a495a3	0.400	0.000	0.545	0.371	0.046	0.499	10	5	11
50f70c6ecb	0.583	0.000	0.167	0.549	0.064	0.162	12	3	18
bab7c2acaf	0.833	0.333	0.750	0.644	0.247	0.715	7	3	20
dd9c1003d5	0.556	0.000	0.250	0.513	0.043	0.210	9	6	4

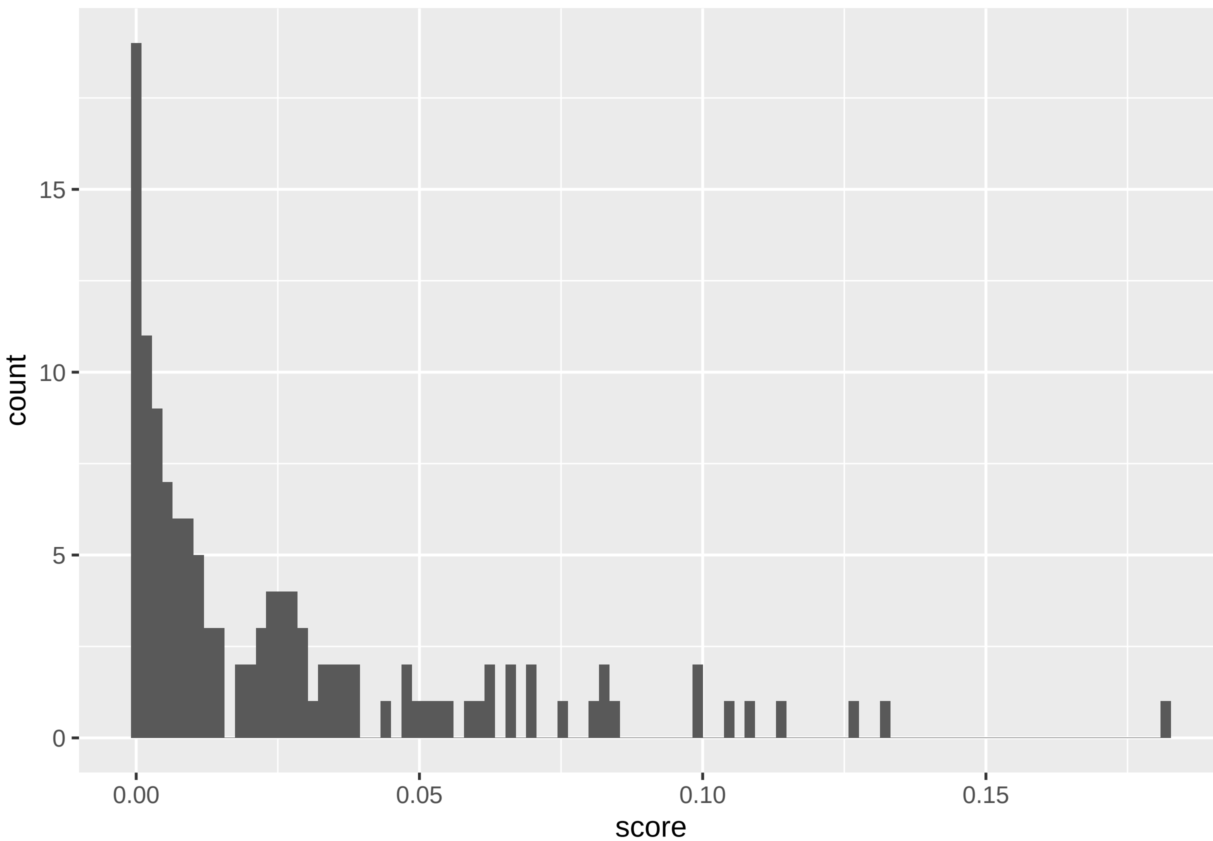


Figure 3: Fairness scores for the officers under consideration.